METHOD AND PRODUCT FOR THE SEQUENCE DETERMINATION OF PEPTIDES USING A MASS SPECROMETER

Brian T. Chait[1], Ronald C. Beavis[1], Rong Wang[1], and Stephen B.H. Kent[2]

1. The Rockefeller University

1230 York Avenue

New York, N.Y. 10021

2. The Scripps Research Institute

10666 North Torrey Pines Road

LaJolla, California 92037

345-201 A
07 891177

FIELD OF THE INVENTION

This invention relates to rapid and efficient methods for sequencing polypeptides (including peptides) and proteins utilizing a mass spectrometer.

Polypeptides are a class of compounds composed of α-amino acid units chemically bonded together with amide linkages with elimination of water. A polypeptide is thus a polymer of α-amino acids which may consist to several thousand amino acid residues. Peptides are similar to polypeptides, except that they are comprised of a lesser number of α-amino acids. There is no clear-out distinction between polypeptides and peptides. For convenience, in this disclosure and claims, the term "polypeptide" will be used to refer generally to peptides and polypeptides.

Proteins are polypeptide chains folded into a defined three dimensional structure. They are complex high polymers containing carbon, hydrogen, nitrogen, and sulfur and are usually comprised of chains of amino acids connected by peptide links. They are similar to peptides, but are usually of a much higher molecular weight.

For a complete understanding of physiological reactions involving proteins it is often necessary to understand their structure. There are a number of facets to the structure of protein. These are the primary structure which is concerned with amino acid sequence in the protein chain and secondary, tertiary and quaternary structure which generally relate to the three dimensional configuration of proteins. This invention is concerned with the determination of primary structure. It provides a facile and accurate procedure for the determination of such structure.

Many procedures have been used over the years to determine the amino acid sequence, i.e. the primary structure, of proteins. At the present time, the best method available for determining the amino acid sequence of a protein chain is the Edman degradation. In this procedure, one amino terminal amino acid residue at a time is removed from a polypeptide to be analyzed. That amino acid is normally identified by reverse phase HPLC, but in recently mass spectrometric procedures have been described for this purpose (Aebersold, R., Bures, E.J., Namchuck, M., Goghari, M.H., Shushan, B., and Covey, T.C., Protein Science 1992, 1, 494). The Edman degradation cycle is repeated for each successive terminal amino acid residue until the complete polypeptide has been degraded. The procedure is tedious and time consuming. Each sequential removal of a terminal amino acid requires 20 to 30 minutes. Hence, with a polypeptide of even moderate length, say for example 50 amino acid residues, a sequence determination may require many hours. The procedure has been automated. The automated machines are available as sequenators, but the procedure still requires an unacceptable amount of time. The procedure is widely employed, but a procedure which required less time and which yielded information about modified or unusual amino acid residues would be very useful to the art. A procedure that can be used on mixtures of proteins would also be very useful to the art.

SUMMARY OF THE INVENTION

The procedure of this invention avoids the time consuming separation and identification steps of the Edman procedure and permits sequencing of polypeptides in one reaction vessel without separation of any reaction products. The essence of the invention is that degradation is conducted by sequencing reaction conditions using a pair of reagents both of which react with a terminal amino acid residue of a polypeptide to be analyzed. One of the reagents forms a first reaction product, a (terminated)

blocked polypeptide chain, which is resistant to all further manipulations. The second reagent reacts with the same residue and forms a second reaction which under appropriate conditions cyclizes and cleaves to give the polypeptide with one less amino acid residue and an additional low molecular weight product which is cleaved from the polypeptide. There is no necessity to isolate, purify and identify the cleaved low MW product. Further cycles of coupling termination are carried out without separation of the blocked (terminated) peptidase chains. After the desired number of cycles, an aliquot (representative sample) of the polypeptide product mixture is taken and subjected to readout in a single mass spectrometry experiment, which determines the masses of all the blocked polypeptide chains. The amino acid sequence is defined by the mass differences between the neighboring peaks in the mass spectrometry experiment. A detailed outline of this approach is given below.

DETAILED DESCRIPTION OF THE INVENTION:

Powerful new approaches to protein covalent structure determination arise from the combination of wet chemical with the newly evolved peptide protein mass spectrometry techniques. The newer mass spectrometry techniques yield useful data from picomole-to-sub-picomole amounts of peptides/proteins. Further the incipient ion-trap technologies promise even better sensitivities, and have already been demonstrated to yield useful spectra in the atamole sample range. In general, both the ion-spray-electrospray and matrix-assisted laser desorption ionization methods mainly generate intact molecular ions. The resolution of the ion-electrospray quadrupole instruments is about 1 in 2,000 and that of the laser desorption time-of-flight instruments about 1 in 400. Both techniques give mass accuracies of about 1 in 10-20,000 (i.e. +/- 0.01% or better). There are proposed modifications of time-of-flight analyzer that may improve the resolution by up to as factor of 10-fold, and thus improve the mass

accuracy of that technique.

Thus, for peptides in the 10-to-50 residue size range now, and the 10-to-100 plus residue range in the future, these techniques yield mass measurements accurate to +/- 0.2 atomic mass units, or better. These capabilities mean that the peptide itself can be analyzed more readily, with greater speed, sensitivity, and precision, than the amino acid derivative released by stepwise degradation techniques such as the Edman degradation. Hence, the underlying imperative of amino acid sequence determination of peptides and proteins has changed. This leads to a new principle of peptide/protein sequence determination, as follows. The ideal approach would now be to take a polypeptide chain and generate a family of "fragments", each differing by a single amino acid, Figure 1.

1-2-3-4-5-6-7-8-9-.....-n-(OH) INTACT STARTING PEPTIDE CHAIN

(X)-1-2-3-4-5-6-7-8-9.....-n-(OH)

(X)-2-3-4-5-6-7-8-9-.....-n-(OH)

(X)-3-4-5-6-7-8-9.....-n-(OH)

(X)-4-5-6-7-8-9.....-n-(OH)

(X)-5-6-7-8-9.....-n-(OH)

(Figure 1)

etc.

where X = H, or any other constant moiety. Typically, X will be a terminating moeity that is resistant to all subsequent manipulations in a degradation experiment.

From such a family of terminated molecular species (a "ragged-end" polypeptide forming a "protein sequencing ladder") the amino acid sequence can be simply read out in a single experimental operation, based on the mass differences between the intact molecular ions (Fig. 2).

Furthermore, the accuracy of the amino acid sequence thus determined is insensitive, over a wide range (5-fold or more), to the amount of each molecular species present in the mixture as to shown in Fig. 2.

Thus the demands on the yields of a chemical degradation reaction are much less stringent and more readily achieved than for the wet chemical stepwise degradation techniques. This approach can be readily adapted to either N- or C-terminal sequences determination.

The protein sequencing ladder (family of "ragged-end" peptides) could be generated from the N-terminal by, for example, carrying out a deliberately inefficient Edman degradation for 10, or 12 (or n) cycles. This could be done in the classic fashion, with the peptide sample trapped (physically or chemically) and then repeating the following cycle (Figure 3):

It would be necessary to"de-tune" the reactions to give a suitable ragged-end

One way of doing this would be as follows: By addition of (say) 5-10% phenyl isocyanates to the PITC, a certain proportion (5-10%) of the polypeptide chains would be irreversibly terminated at each cycle, thus generating the desired "ragged-end". The amount of terminating reagent can be empirically adjusted, and can be increased or decreased as a function of cycle number, to give the desired level of termination. An aliquot of the product mixture of blocked polypeptide is then read out by mass spectrometry. The low MW products of the degradation chemistry would not interfere in the mass range of interest for most peptides. Reagent pairs (degradation reagent/blocking reagent) need not be chemically related.

However, there is a completely new possibility which presents itself because of the different underlying principle of the sequencing experiment: instead of intervening between the coupling and cleavage steps for washes to remove reagents, one could keep the Edman (PITC) or other reagent present at all times and merely cycle the reaction conditions.

For example, use an excess of PITC with an added preparation of suitable terminating reagent, and cycle the pH of the reaction mixture between strongly acidic and basic, alternating several minutes at each pH. No washes or separations between steps, all products accumulate in the solution. The PITC may be stable to both sets of conditions, or it may be necessary to add more to each "basic pH" step.

This constitutes a one-pot cyclic degradation. There are no handling steps, hence no sample loss. Volatilize reagents can be used, so that sample concentration is straightforward, if required.

A constant 10% termination at each cycle would yield a ragged-end polypeptide mixture of the following molar composition, after 10 cycles:

	Mole Fraction
(X)-1-2-3-4-5-6-7-8-9-10-11-12-.....-n-(OH)	0.10
(X)-2-3-4-5-6-7-8-9-10-11-12-.....-n-(OH)	0.09
(X)-3-4-5-6-7-8-9-10-11-12-.....-n-(OH)	0.08
(X)-4-5-6-7-8-9-10-11-12-.....-n-(OH)	0.07
(X)-5-6-7-8-9-10-11-12-.....-n-(OH)	0.07

(X)-6-7-8-9-10-11-12-.....-n-(OH)	0.06
(X)-7-8-9-10-11-12-.....-n-(OH)	0.05
(X)-8-9-10-11-12-.....-n-(OH)	0.05
(X)-9-10-11-12-.....-n-(OH)	0.04
(X)-10-11-12-.....-n-(OH)	0.04
(X)-11-12-.....-n-(OH)	0.35

The mass of the terminating agent moiety, X, is irrelevant, as the sequence is determined from the mass differences. The mass spectrum of this mixture is shown in Fig. 2. Thus, there would be approximately equal abundance of the sequence-determining molecular ions in the protein sequencing ladder as read-out by mass spectrometry.

This principle of "one-pot cyclic degradation with mass spectrometry read out" represents a completely new approach to protein sequence determination. The advantages are obvious; speed, convenience, sample recovery, & hence sensitivity, together with relative insensitivity to reaction yields. Relatively unsophisticated and inexpensive mass spectrometric instrumentation (e.g. time-of-flight; single quadrupole; etc) can be used.

In this way it may be routinely possible to obtain 10 plus residues of sequence from one picomole or less of a polypeptide chain, in less than 30 minutes, including the

cyclic degradation, mass spectrometry, and interpretation. Cyclic degradation of a large number of separate peptide samples can be run simultaneously, then, the mass spectrometry read out becomes limiting; say, 10-100 plus residues per minute (conservatively).

It will be apparent to one skilled in the art that the processes described may be readily automated i.e., carried out for example in microtiter plates, using an x,y,z chemical robot. Furthermore, the determination of amino acid sequence from mass spectrometric data obtained from the protein sequencing ladders is readily carried out by simple computer algorithms.

This approach converts the inelegant and time-consuming linear, stepwise chemical approach or the complex, difficult to interpret mass spec/mass spec (tandem mass spectrometric) approach to a simple, highly-redundant parallel-readout approach. This simplifies the method, speeds it up enormously, and reduces the chance of error.

The overall process is as follows:

PROTEIN SAMPLE

[few tens of picomoles (now)
few picomoles or less (later)]

↓ (enzymatic digest)
(e.g. tryptic)

fragments

↓ separate by HPCL

collection of separated peptides*

↓ parallel cyclic degradations

collection of ragged end peptides

↓ mass spectrometry readout

extensive sequence data.

*It may not be necessary to completely separate all the peptides from one another because the present sequencing techniques can be effectively used to sequence mixtures of peptides.

N- or C-terminal amino acid sequence data may be obtained by direct treatment (cyclic degradation) of intact proteins.

Finally, as an alternative, it is possible to obtain extensive sequence data from all the proteins and polypeptides on a 1- or 2-D gel by electroblotting, followed by "in situ" cyclic degradation (using, for example gas phase reagents), and matrix-assisted laser desorption T.O.F. readout in a position-dependent fashion. It is a particular feature of the invention that it is not limited to analysis of one peptide. Mixtures of peptides can be analyzed simultaneously in one reaction vessel. Each will give a separate curve as shown in an idealized form in Fig. 4. In this figure the molecular weights of the original components of the mixture differ by 500 mass units. Each of the separate curves can be analyzed in the same manner as Fig.3. Fig. 4 shows the type of curve obtained when there is appreciable overlapping in molecular weight amongst the polypeptides to be sequenced.

The process of the present invention has certain inherent limitations. First Leu/Ile have the same mass, but, the difference between them may be obtained from cDNA sequencing. They are highly degenerate codons, so they can be accommodated by inosine substitution in DNA probes/primers for isolation/identification of the corresponding gene. This limitation will have little impact on the practical utility of the method.

Secondly, several amino acids differ by only 1 amu which places stringent requirements on mass accuracy. However, we need to determine mass differences between adjacent peaks - this can be done much more accurately than the absolute mass of the peptides. Hence, in practice this will not be a significant limitation. Third, peptide/protein samples which are blocked at the N- or C-terminal will not be degraded.

This can be circumvented by chemical or enzymatic fragmentation of the blocked polypeptide chain to yield unblocked segments.

Variations on the above process, yielding similarly useful amino acid sequence information, are as follows:

1. As an alternative to stepwise degradation of the polypeptide chain by coupling under one set of conditions (e.g. base) followed by cleavage/cyclization under a second set of conditions (e.g. acid), a reagent can be used which couples to a terminal amino acid of a polypeptide chain and cyclizes/cleaves under a single set of conditions. e.g. base. In the presence of a suitable terminating agent, such a reaction will yield the desired protein sequencing ladder (ragged end [blocked] polypeptide mixture) in a single operation, merely by treating the polypeptide with the two reagents under appropriate conditions. Such reagents are known, but have previously been regarded as poor candidates for use in sequencing because of their unsuitability for controlled, stepwise-degradation. Their properties are ideal for the current application.

2. Termination-by-side-reaction. Use of a single reagent yields two products - a blocked polypeptide chain and a polypeptide chain one residue shorter.

3. Determination of the amino acid sequence of a target polypeptide chain which has been prepared by stepwise solid phase peptide synthesis. Stepwise solid phase peptide synthesis involves the assembly of a protected peptide chain by repetition of a series of chemical steps (the "synthetic/cycle") which result in the addition of one amino acid residue to a polymer-bound peptide chain. The target polypeptide chain is built up one residue at a time, usually from the C-terminal, by repetition of the synthetic cycle. Peptide-resin samples can be taken after each cycle. Mixing approximately equal

amounts of all samples obtained in the course of a synthesis yields all possible lengths of peptide-resin. Cleavage/deprotection of such a mixture of peptide-resin samples yield a mixture of free polypeptide chains of all possible lengths, viz

AA_7

AA_6-AA_7

AA_5-AA_6-AA_7 for a 7 residue sequence

AA_4-AA_5-AA_6-AA_7

etc.

Plus minor amounts of by-products - typically less than a few percent for each peptide chain length.

Readout of an aliquot of the mixture, by e.g. matrix-assisted laser desorption time-of-flight mass spectrometry, yields a sequence-defining set of molecular ions, as described above. Irreversibly terminated by-products formed in the course of the synthesis are present in all subsequent samples and are thus amplified. This constitutes a uniquely effective way of detecting such side reactions.

EXAMPLES:

Example 1:

The above-described procedure was tested on a 14-residue peptide, [Glu^1]-Fibrinopeptide B, with the sequence (one letter code) EGVNDNEEGFFSAR. A modification of the original manual Edman degradation procedure[1] was employed. To generate the **desired** peptide ladder, a mixture of phenylisothiocyanate (PITC) and phenylisocyanate (PIC), with ratio 20:1, was used in the step-wise degradation reaction. Here, PITC is used as coupling agent and PIC is used as terminating agent.

Materials and chemicals. [Glu^1]-Fibrinopeptide B was purchased from Sigma Chemical Co. and used with no further purification. Phenylisothiocyanate (PITC, sequanal grade), Pyridine (sequanal grade), and Trifluoroacetic acid (TFA, HPLC/spectro grade) were obtained from PIERCE. 99% Trimethylamine (TMA), Phenylisocyanate (PIC), Ethyl Acetate (EA, HPLC/spectro grade), and 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) were obtained from Aldrich. Heptane was obtained from Burdick & Jackson. 25% Trimethylamine (TMA) was prepared by dissolving 3.80 g of 99% TMA in 11.40 g water.

Edman degradation. All of reaction cycles were carried out in a single 1.5 ml Eppendorf microcentrifuge vial under a stream of dry N_2. The coupling step was carried out as follows: Peptide (200 pmoles to 10 nmoles) was dissolved in 20 µl of 25% TMA:Pyridine (1:1). 20 µl of coupling agent containing PITC:PIC:Pyridine:HFIP (20:1:76:4, v/v) was added into the reaction vial. The coupling reaction was allowed to proceed at 50 °C for 3 minutes. The coupling agent and side-reaction products

were extracted as follows: 400 µl of heptane:EA (10:1, v/v) was added into the reaction vial, gently vortexed and centrifuged to clear phases. The upper phase was aspirated and discarded. The above wash procedure was repeated once and followed by twice washing with heptane:EA (2:1, v/v). The remaining solution in the reaction vial was lyophilized in a Speed-Vac centrifuge. The cleavage step was carried out by adding 20 µl of anhydrous TFA to the dry residue in the reaction vial and allowed to react at 50 °C for 5 minutes, followed by lyophilization. The above-described coupling-cleavage cycle were repeated a further six times.

Matrix-assisted laser desorption mass spectrometric (MALDMS) analysis. To measure the molecular weight of the starting peptide, the peptide was first dissolved in 100 µl water and 2 µl of the resulting solution was then transferred into 10 µl of water as sample solution for MALDMS analysis. The remaining peptide solution was lyophilized for Edman degradation. Similarly, after seven cycles of modified Edman degradation, the peptide mixture was dissolved in water and 4% of peptide was transferred into 10 µl water for MALDMS analysis. 1 µl of each of the above sample solution was individually mixed with 9 µl of matrix solution containing 5 g/l of α-cyano-4-hydroxycinnamic acid (4HCCA) in 0.1% TFA:Acetonitrile (2:1, v/v).[2] 0.5 µl of the resulting mixture was applied to the probe of a laser desorption time-of-flight mass spectrometer. Mass spectra were obtained on a laser desorption time-of-flight mass spectrometer constructed at The Rockefeller University and described previously.[3]

Peptide sequence read-out. Positive ion MALDMS spectra of [Glu^1]-Fibrinopeptide B is shown in Figure 1.* A protonated molecular ion $[M+H]^+$ was

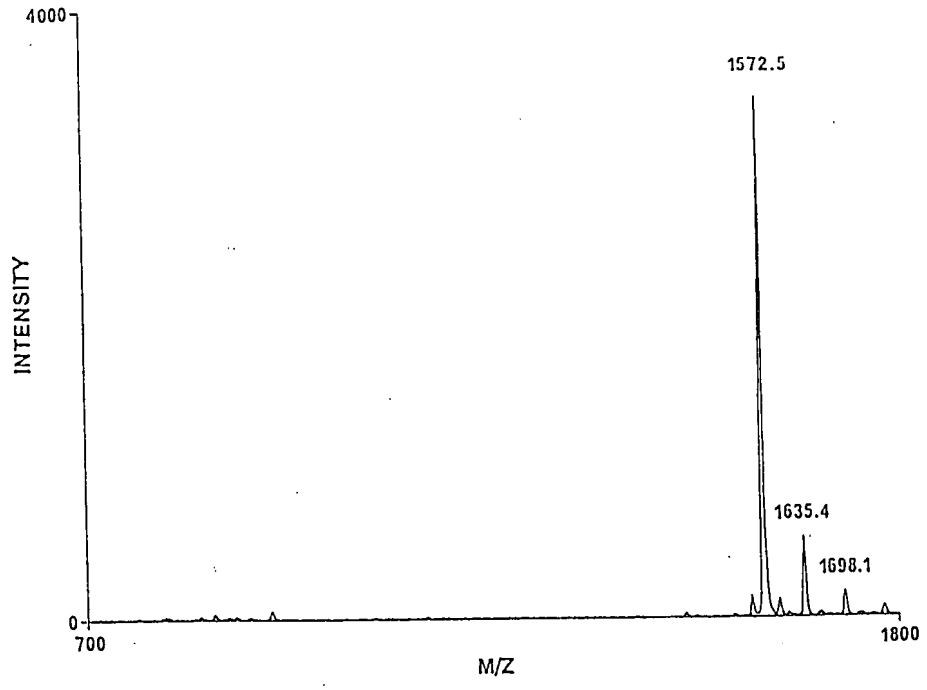

Figure 1.* Positive ion matrix-assisted laser desorption mass spectrum of [Glu[1]]-Fibrinopeptide B.

observed at m/z 1572.5 (calculated value is 1571.8).

As described previously, A peptide ladder of [Glu[1]]-Fibrinopeptide B was formed after seven cycles of modified Edman degradation. Its positive ion MALDMS spectrum is shown in Figure 2.* Each of the peaks in the spectrum represents a related phenylcarbomyl peptide derivative in the peptide ladder (except a few peaks which will discussed later). The amino acid sequence can be easily read-out from the mass difference of adjacent two peaks. For instance, the mass differences are 129.1, 56.9, and 99.2 between peaks at m/z 1690.9 and 1561.8, peaks at m/z 1561.8 and 1504.9 and peaks at m/z 1504.9 and 1405.7. Which correspond to glutamic acid (ca. 129.12), glycine (ca. 57.05) and valine (ca. 99.13) residues, respectively. Seven amino acids read-out in this manner are denoted

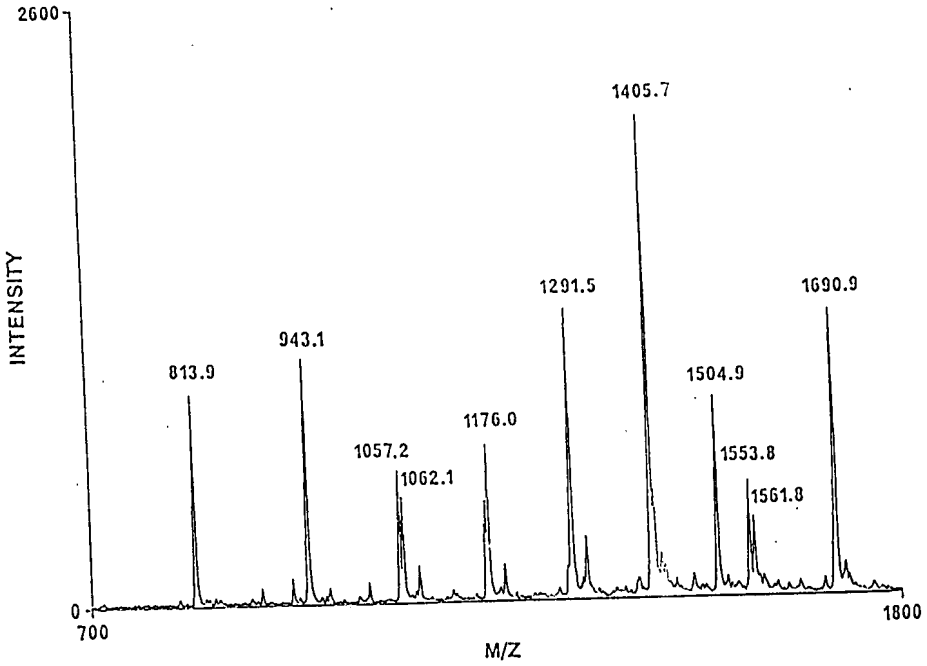

Figure 2. Positive ion matrix-assisted laser desorption mass spectrum of [Glu[1]]-Fibrinopeptide B after 7 cycles of modified Edman degradation.

under the x-axial in the spectrum (Figure 2). One of paired peaks gives mass difference 119.0 (1062.1 - 943.1) is resulted from the PIC. In another word, these two peaks are represent one piece of peptide with or without phenylcarbomul group. Peak at m/z 1553.8 corresponds partially blocked peptide with pyroglutamic acid at the N-terminus.

References:

1. G.E. Tarr (1977), in *Methods Enzymology* 47, 355.

2. R.C. Beavis T. Chaudhary and B.T. Chait (1992), *Organic Mass Spectrom.* **27**, 156

3. R.C. Beavis and B.T. Chait (1989), *Rapid Commun. Mass Spectrom.* **3**, 233.

Example #2

Stepwise solid phase synthesis of the 99 amino acid residue polypeptide chain corresponding to the monomer of the HIV-1 protease (SF2 isolate):
PQITLWQRPLVTIRIGGQLKEALLDTGADDTVLEEMNLPGKWKPKMIGGIGGFIKVRQYD
QIPVEI(Aba)GHKAIGTVLVGPTPVNIIGRNLLTQIG(Aba)TLNF
[where Aba = α-amino-*n*-butyric acid]
Highly optimized Boc-chemistry instrument-assisted stepwise assembly of the protected peptide chain was carried out on a resin support, according to S.B.H. Kent, *Annual Rev. Biochem.* **57**, 957-984 (1988). Samples (3-8mg, ~1μmole each) were taken after each cycle of amino acid addition. The protected peptide-resin samples were mixed in three batches of consecutive samples: (number corresponds to the amino acid after which sample was taken, i.e. residue #)99-67; 66-33; 32-1. The mixed batches of peptide-resin were deprotected and cleaved with HF (1 hour, at O°C, plus 5%cresol/5%thiocresol). The products were precipitated with diethyl ether, and dissolved in acetic acid-water (50/50%, v/v), then lyophilized.

1 μl of the peptide mixture (10 μM per peptide component) was added to 9 μl of 4-hydroxy-α cyanocinnamic acid in a 1:2 (v/v) ratio of 30% acetonitrile/0.1% aqueous trifluoroacetic acid. 1/2 μl of the resulting mixture was applied to the mass spectrometer probe and inserted into the instrument (R.C. Beavis and B.T. Chait, Rapid Commun. Mass Spectrom. 3 (1989) 233). The spectra shown below are the result of 100 laser shots at a rate of 2.5 laser shots/second. Figure A shows the mass spectrum obtained from the mixture 99-67 (actually 99-n, where n = 99,67). The labels on the peaks, n, refer to the peptides having residues 99-n. Table 1 shows the measured masses of a selection of these peaks and compares them with the known sequences of the peptides. The agreements are sufficiently close to allow confirmation of the correctness of the synthesis.

Figure B shows the mixture (66-33) [more strictly (99-n, where n = 66,33)] and Table II corresponding mass data obtained from the spectrum.

The sequence of the assembled polypeptide chain can be read out in a straightforward fashion from the mass differences between consecutive peaks in the mass spectra of the peptide mixture. This confirmed the sequence of amino acids in the peptide chain actually synthesized.

Peak	Measured MM	Calculated MM	Δ
71	2977.9	2977.5	+0.4
72	2906.5	2906.5	0. 0
73	2793.1	2793.3	−0.2
74	2736.3	2736.3	0. 0
75	2635.2	2635.2	0. 0
76	2536.0	2536.0	0. 0
77	2422.9	2422.9	0. 0
78	2323.8	2323.7	+0.1
79	2266.7	2266.7	0. 0
80	2169.5	2169.6	−0.1
81	2068.4	2068.4	0. 0
82	1971.3	1971.3	0. 0
83	1872.1	1872.2	−0.1
84	1758.3	1758.1	+0.2
85	1644.9	1644.9	0. 0
86	1531.8	1531.8	0. 0

TABLE II

Peak	Measured MM	Calculated MM	Δ
44	5926.1	5926.1	0. 0
43	6054.3	6054.2	+0.1
42	6240.7	6240.4	+0.3
41	6368.8	6368.6	+0.2
40	6424.9	6425.7	−0.8
39	6522.8	6522.8	0. 0
38	6635.2	6635.9	−0.7
37	6750.2	6750.1	+0.1
36	6881.0	6881.2	−0.2
35	7010.5	7010.4	+0.1
34	7140.2	7139.5	+0.7
33	7253.5	7253.6	−0.1

In addition, terminated by-products (where the peptide chain has become blocked and does not grow any more) are present in every subsequently taken peptide-resin sample. Thus, there is an amplification factor equal to the number of resin samples in the batch after the point of termination. This can be seen in Figure (samples #66-33) which contains a peak at 3339.0. This corresponds to the peptide 71-99, 3242.9 (N-terminal His71) plus 96.1 dalton. This is the trifluoroacetyl-peptide, Nα-Tfa-(71-99). The ratio of the amount of this component to the average amount of the other components is about 2:1. There were 34 samples combined in this sample. Thus, the terminated byproduct Nα-Tfa-(71-99) had occured at a level of ~5mol%. This side reaction, specific to the N-terminal His-peptide chain, has not previously been reported. This illustrates the important sensitivity advantage provided by this amplification effect in detecting terminated peptides. Such byproducts are not readily detected by any other means.

Figure C shows the low mass region of the (66-37) mixture showing a number of peaks corresponding to side reaction products. The identification of one major product with mass 3339.0 is made in Figure D.

FIGURE D 29

891177

IDENTIFICATION OF SIDE-REACTION PRODUCT

MEASURED MM OF MAJOR SIDE-PRODUCT = 3339.0

CALCULATED MM OF PEPTIDE (99-69) = 3242.9

DIFFERENCE = 96.1

HIS-69 APPEARS TO BE TRIFLUOROACETYLATED

Example #3.

An alternative way of generating an amino acid sequence-defining collection of all possible length peptides derived from a single polypeptide chain is as follows. An mRNA coding for the polypeptide chain of interest is converted to the cDNA and is amplified by standard methods. This cDNA is then used in *in vitro* transcription to create a larger amount of the mRNA. *In vitro* translation of the mRNA is carried out by standard methods (Ref. C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, / Science 244, 182-188 (1989). These procedures yield multiple tens of µg (i.e. nanomole) quantities of protein. In the present case, the antibiotic puromycin is added to the *in vitro* translation system. This antibiotic mimics an aminoacyl-tRNA, adds to the carboxyl terminal of the growing peptide chain to form a covalent adduct and causes premature release of the peptide chain from the ribosome-mRNA complex. A suitable level of the antibiotic is used to create a low (few % per amino acid residue) approximately uniform level of prematurely released carboxyl terminal-modified polypeptide chains. The resulting mixture of polypeptide chains will represent all possible lengths from the N-terminal (aa1-puromycin, aa1-aa2-puromycin, aa1-aa2-aa3-puromycin, aa1-aa2-aa3-aa4-puromycin, aa1-aa2-aa3-aa4-aa5-puromycin, etc). Matrix-assisted laser desorption mass spectrometric read-out of such a mixture will give the pattern of peaks shown in Figure . Individual peptide chain components of each length will be present in picomole to multiple-picomole amounts. The mass differences between consecutive peaks defines the amino acid sequence of the N-terminal of the polypeptide coded for by the original mRNA. A further advantage of the puromycin covalent adduct of the terminated peptide chains is that the added 472 daltons boosts the mass of the short, N-terminal defining peptides above the range where the matrix components interfere with identification of the peptides.

From time to time it may be desirable to determine which codon is being used for a particular amino acid, and thus surmount the degeneracy of the genetic code. This can be accomplished by using a tRNA corresponding to the triplet codon in question, misloaded (Noren, et al., loc.sit.) with a unique mass amino acid, or with a modified amino acid containing a marker atom such as Br. It is possible to read through termination codons (Noren et al., loc sit) using suitable suppressor tRNA molecules acylated with unique mass amino acids. This will result in unique mass differences in the polypeptide read out experiment, uniquely identifying the corresponding codons at the nucleic acid level. In this way, nucleic acid sequences can be read as peptide sequences. This has a two-fold advantage: 1. three nucleotides are represented as a single amino acid; 2.

an amino acid is on average only about 35% the mass of a nucleotide. These two effects result in an almost 10-fold compression of the mass range needed to represent a given nucleic acid sequence.

mRNA ———————————

↓ in vitro translation } in the presence of puromycin (Δmass 472)

aa_1-puromycin] Δmass = aa_2
aa_1-aa_2-puromycin] Δmass = aa_3
aa_1-aa_2-aa_3-puromycin
etc. etc.

Mass Spectrometry Read-Out: